Detecting syntactic and semantic anomalies in schizophrenia

Andrea Moro a, Valentina Bambini a,*, Marta Bosia b,c, Simona Anselmetti c, Roberta Riccaboni c, Stefano F. Cappa a,d, Enrico Smeraldi a,c, Roberto Cavallaro c

a Center for Neurocognition, Epistemology and theoretical Syntax (NEtS), Institute for Advanced Study (IUSS), Pavia, Italy
b Università Vita Salute San Raffaele, Milan Italy
c Department of Clinical Neurosciences, IRCCS San Raffaele Scientific Institute, Milan, Italy
d Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy

A R T I C L E   I N F O

Article info
Received 17 June 2015
Received in revised form 18 October 2015
Accepted 25 October 2015
Available online 28 October 2015

Keywords:
Language
Schizophrenia
Syntax
Semantics
Anomaly
Grammaticality judgments

A B S T R A C T

One of the major challenges in the study of language in schizophrenia is to identify specific levels of the linguistic structure that might be selectively impaired. While historically a main semantic deficit has been widely claimed, results are mixed, with also evidence of syntactic impairment. This might be due to heterogeneity in materials and paradigms across studies, which often do not allow to tap into single linguistic components. Moreover, the interaction between linguistic and neurocognitive deficits is still unclear. In this study, we concentrated on syntactic and semantic knowledge. We employed an anomaly detection task including short and long sentences with either syntactic errors violating the principles of Universal Grammar, or a novel form of semantic errors, resulting from a contradiction in the computation of the whole sentence meaning. Fifty-eight patients with diagnosis of schizophrenia were compared to 30 healthy subjects. Results showed that, in patients, only the ability to identify syntactic anomaly, both in short and long sentences, was impaired. This result cannot be explained by working memory abilities or psychopathological features. These findings suggest the presence of an impairment of syntactic knowledge in schizophrenia, at least partially independent of the cognitive and psychopathological profile. On the contrary, we cannot conclude that there is a semantic impairment, at least in terms of compositional semantics abilities.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Language impairments have been reported by Bleuler since his first description of schizophrenia, and even earlier by Kraepelin, as features of dementia praecox. In the last edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), severely disorganized speech, substantially impairing effective communication, is classified as one of the diagnostic criteria for schizophrenia. Despite their acknowledgment as important features of the illness, language disturbances are generally considered as a reflection of the underlying disorders of thinking, rather than primary impairments. Clinical descriptions typically encompass derailment or loose associations, tangentiality, incoherence or “word salad” and neologisms, attributed to disorganized thinking, as well as mutism, observed in catatonic behavior, and diminished speech output and prosodic modulation, as expression of negative symptoms.

Modern approaches have documented deficits in language processing, involving both comprehension and production (DeLisi, 2001). These impairments at the behavioral level are also related to functional and anatomical alterations in the brain networks for language (Catani et al., 2011; Leroux et al., 2013; Benetti et al., 2015) and altered processing dynamics (Kuperberg, 2010b). More recently, also in response to advancements in linguistics and neuroscience of language, the challenge has become to identify specific components of the linguistic structure where patients with schizophrenia might exhibit selective deficits (Covington et al., 2005). Since the nineties, one of the main lines of investigations on the neurobiology of language has become to disentangle the relative contribution of speech production and perception, syntax, semantics and more recently pragmatics to language impairment in schizophrenia. However, evidence is mixed for all components, leaving the description of the linguistic profile of patients with schizophrenia still a matter of debate.

It has long been claimed that speech produced by patients with schizophrenia is characterized by “preoccupation with too many of the semantic features of words” (Chaika 1974, cited in Kuperberg and Caplan, 2003). A large number of studies have focused on the semantic component, investigating difficulties in meaning processing through a variety of paradigms. Among these, the most
common are priming tasks, which point to hyperpriming effects and disorganized semantic storage (Spitzer et al., 1994; Rossell and David, 2006). Evidence of disorganization of the semantic system comes also from studies employing vocabulary and naming tests (Goldberg et al., 1998). These findings, however, are far from consistent. Other studies reports that, depending on the stimulus onset asynchrony, patients might exhibit hypopriming effects or even performance comparable to healthy controls (Barch et al., 1996). Moreover, many studies suggest that the problem might involve high-order semantics, for instance affecting long distance associations while sparing naming (Barrera et al., 2005), lexical co-occurrences and sequences of discourse (Elvevag et al. 2007; Hella et al., 2013). Other studies have considered most complex aspects of meaning interpretation, such as the comprehension of figurative language, among which idiomatic expressions (Schettino et al., 2010), metaphor and irony (Langdon et al., 2002; Champagne-Lavau and Stip, 2010). A deficit in the ability to derive and to produce context-appropriate meanings seems to be a common feature of schizophrenia, suggesting that the semantic impairment might rather be ascribed to high-order semantics or even pragmatics and discourse.

As for the syntactic component, the historical view described the syntax of schizophrenic speech as normal, even when the organization of discourse is completely broken down, as in this example of word salad: If we need soap when you can jump into a pool of water, and then when you go to buy your gasoline, my folks always thought they should get pop, but the best thing is to get motor oil... (Andreasen, 1979, cited in Convinton et al., 2005). Also for sentence comprehension, old studies employing the embedded click paradigm reported preserved sensitivity to syntactic boundaries (Rochester et al., 1973; Carpenter, 1976; Grove and Andreasen, 1985). However, more recent literature depicted a different scenario. A series of studies on production indicate the presence of different syntactic errors, linked to negative symptoms (Thomas et al., 1987), more severe in chronic patients (Thomas et al., 1990), although not progressing during a follow-up study (King et al., 1990). Imaging data support the neuropsychological findings, documenting significant differences in the pattern of correlation of brain activity with sentence complexity during production between controls and patients affected by schizophrenia (Kircher et al., 2005). Syntactic comprehension has also been described as compromised. Morice and McNicol (1985) developed a complex version of the Token Test, including sentences with embedding, and found a reduced performance, correlated with simplified syntax (less embedding) in production, in patients with schizophrenia. Condray et al. (1996) used a different test, in which patients had to answer questions about the meaning of syntactically complex sentences. Reduced comprehension accuracy was found in patients as compared to controls. Using a similar test, Bagn er et al. (2003) found that increasing sentence length and complexity had a greater impact on language comprehension performance in patients as compared to controls. Other studies employed the picture matching task and also observed reduced access to syntactic structure (Lelekov et al., 2000; Tavano et al., 2008).

A major topic of discussion is whether language impairment can be considered specific or it is the result of psychopathological condition and cognitive deficit (Kuperberg, 2010a; Kuperberg and Heckers, 2000). Among symptoms, the role of thought disorder has been widely discussed in literature, and related especially to semantics and discourse (Rodriguez-Ferrera et al., 2001; McKenna and Oh, 2005). As for cognitive functions, a crucial issue in the explanation of defective language comprehension is the role of working memory deficits. Several of the studies mentioned above reported a significant positive correlation between the patients’ working memory abilities and sentence comprehension accuracy (e.g., Condray et al. 1996; Bagn er et al. 2003). According to Kuperberg (2010b), however, it is possible that these correlations reflect unmeasured variables, or that working memory affects difficulty in assigning thematic roles or plausibility judgments rather than syntactic abilities per se.

In reviewing the literature, another important aspect to note is that most of the paradigms tapped onto more than one linguistic component, and do not really allow to disentangle the different linguistic aspects as defined in linguistic theory. Semantic tasks often involved a combination of lexical and world knowledge, pragmatics and discourse organization. Conversely, syntactic tasks are often intertwined with morphology or with semantics, as in cases where agreement and thematic roles are involved. Besides, analysis of production is not sufficient to detect possible underlying impairments. This issue makes research on language in schizophrenia restrained to descriptions that are scarcely more precise than the observations of word salad and tangential speech reported in the early literature.

A useful approach to investigate specific linguistic components is the anomaly detection paradigm, as anomalies can be constructed in such a way that they selectively involve specific levels of the linguistic structure. The detection of violations, at different levels, has been extensively studied in healthy subjects with behavioral paradigms (Bambini et al., 2013), with evoked potentials (Hahne and Friederici, 2002) and fMRI (Moro et al., 2001), as well as in aphasic patients (Linebarger et al., 1983; Grodzinsky and Finkel, 1998; Wilson and Saygin, 2004). In patients, this method offers the additional advantage of allowing to assess aspects of grammatical knowledge that would not emerge in spontaneous production or in the comprehension of grammatical sentences. Defective detection of syntactic anomalies was also observed in patients affected by corticobasal degeneration without clinical evidence of aphasia, independent of impaired grammatical comprehension (Cotelli et al., 2007).

In schizophrenia, the anomaly paradigm has been applied to explore the role of formal thought disorder. In online judgment, though-disordered patients were insensitive to anomalies, irrespectively of the type of violation (either syntactic, semantic or pragmatic) (Kuperberg et al., 1998, 2000). In another study, equal N=400 components were elicited in patients and controls in response to semantic mismatch, while for syntactic errors equal response in earlier time windows (ELAN) contrasted with no response in later time windows (P600) in patients as compared to controls, suggesting intact phrase structure construction but impaired integration (Ruchsow et al., 2003). Moreover, patients showed reduced P600 amplitude as compared to controls in response to both syntactic and animacy-semantic violations (Kuperberg et al., 2006). Overall, these findings point to some form of linguistic impairment but do not offer conclusive evidence, as the types of anomalies employed in the literature showed an amalgam of morphosyntactic, lexical, semantic and pragmatic domains, differently combining agreement errors, subcategory errors at the syntax/semantic interface and world knowledge and discourse violations.

In this study, we applied the anomaly detection paradigm to a sample of patients with schizophrenia and healthy controls to assess syntactic and semantic knowledge and disentangle possible specific impairments. We devoted special care to the construction of the materials, in order to create violations that selectively tap into either syntactic or semantic operations. As for syntax, we employed syntactic errors based on Universal Grammar rules, thus specifically not intertwined with morphological operations. As for semantics, instead, we aimed at tapping into semantic competence, excluding the influence of other processes at the level of world knowledge and discourse. For this purpose, we constructed a set of anomalies based on semantic errors that require both a full computation of the meaning of the sentence and the recognition of semantic “contradiction” within the same lexical domain, as
opposed to a local clash of two lexical items pertaining to two
standardly incompatible semantic fields. In this way, we aimed at
measuring genuine semantic competence, as resulting from com-
positional semantic abilities, independently of lexical competence
and world knowledge. To assess the impact of working memory
load, both long and short sentences were employed. We also
aimed at exploring the role of psychopathological and neu-
ropsychological factors in determining syntactic and semantic
judgment with a battery of ad-hoc tests.

Based on the existing evidence, we hypothesized that patients
with schizophrenia would show impairment in the detection of
syntactic violations independent of working memory functioning
and formal thought disorder. As for semantics, although the type of
anomaly is novel and predictions cannot be based on the existing
literature, we hypothesized normal performance, as no involvement
of high-level semantics, discourse and pragmatics is required.

2. Materials and methods

2.1. Subjects

Fifty-eight patients with DSM-IV-TR (American Psychiatric Asso-
ciation, 2000) diagnosis of chronic schizophrenia, all subtypes, age 18
through 65 years (32 males and 26 females), were recruited from the
Department of Clinical Neurosciences, San Raffaele Scientific Institute,
Milan. Diagnoses were made by staff psychiatrists with clinical in-
terviews. Exclusion criteria were: history of substance dependence or
abuse, comorbid diagnosis on Axis II, epilepsy or any other major
neurological illness or perinatal trauma. All patients included in the
study were treated with a stable dose of antipsychotic monotherapy
for at least 6 months and found to be responsive (30% or more re-
duction of PANSS Total Score). Thirty control subjects, age 18 through
65 years (11 males and 19 females), were recruited from hospital staff
and general population and were screened for psychiatric diag-
nosis and family history on the basis of a clinical interview. All sub-
jects were native speakers of Italian and no bilingual participants
were included in the study. After a complete description of the study,
informed consent to participation was obtained. The protocol was
approved by the local Ethical Committee and followed the principles
of the Declaration of Helsinki.

2.2. Psychopathological assessment

The subgroup of patients was assessed with the following scales, administered by a trained psychiatrist:

2.2.1. Positive and Negative Syndrome Scale for Schizophrenia
(PANSS)

The PANSS (Kay et al., 1987) is a standardized measurement for
typological and dimensional symptoms assessment. It includes 30
items that provide balanced representation of positive and nega-
tive symptoms and gauges their relationship to one another and to
global psychopathology. It constitutes of three subscales (Positive,
Negative, General), assessing positive symptoms, negative symp-
toms and general psychopathology, respectively. A global measure
of illness severity can be derived from the sum of the three sub-
scales (Total PANSS). The PANSS was administered by trained
psychiatrists, through a semi-structured interview, lasting ap-
proximately 30–45 min.

2.2.2. Thought and Language Index (TLI)

The TLI (Liddle, 2002) is a scale measuring the form of thought
and language in psychotic illness. Speech samples are elicited
using a standardized procedure and specific types of aberrant
thought and language are quantified according to a standardized
protocol. The categories of abnormality defined in the TLI are as
follows: poverty of speech, weakening of goals, perseveration,
looseness, peculiar word usage, peculiar sentence construction,
non-logical reasoning, distractibility

2.3. Neuropsychological assessment

All subjects were assessed with the following tasks, adminis-
tered by a trained psychologist:

2.3.1. National Adult Reading Task (NART) adapted to Italian as Test
di Intelligenza Breve (TIB)

The NART (Nelson, 1982), elaborated in its Italian version
(Sartori et al., 1997), is used to estimate premorbid mental ability.
It consists in a rapid reading of 54 Italian words: 34 low frequency
words and 20 high frequency words. The 34 low frequency words
contain both regular and irregular stressed words. The evaluation
is given by calculating the efficiency in assigning the proper stress
to each low frequency word. The total number of mistakes of
reading defines the TIB error score. Estimated IQ scores are then
calculated through the regression of equations taking into account
sex, age and educational level.

2.3.2. Working memory test (N-Back)

Working memory capacity is studied by using a parametric ‘n-
back’ working memory task (Braver et al., 1997; Goldberg et al.,
2003), a computerized task involving increasing cognitive load (0-
back, 1-back and 2-back). Numbers appear randomly on the four
corners of a diamond. Each number has its own unique and fixed
position. The subject is asked to give a response by pressing a
button box with buttons arrayed in the same configuration than the
diamonds. During no-back condition, the subject presses the
button corresponding to the number seen on the screen. During the
one- and two-back conditions, the subject was required to
respond to the number seen one or two stimuli before, respecti-
vely. The percentages of correct responses in the one- and two-
back conditions were used as outcome scores.

2.3.3. Brief Assessment of Cognition in Schizophrenia (BACS)

The subgroup of patients was also assessed with the Brief As-
sessment of Cognition in Schizophrenia (BACS) adapted to Italian
(Anselmetti et al., 2008), administered by a trained psychologist.
The BACS (Keefe et al., 2004) is a brief battery assessing the main
neurocognitive functions that are usually impaired in patients
with schizophrenia. The completion of the entire battery takes
about 30 minutes, depending on the patient’s performance, and it
includes the following tasks:

- **Verbal memory and learning**: patients are presented with 15
  words and are asked to remind as many of these words as possi-
  ble, then they are asked to recall them in whatever order.
  The whole procedure is repeated 5 times. Measures: number of
  recalled words at the 5th trial.
- **Working memory**, evaluated with a sequence of numbers task.
  Patients are read groups of numbers (i.e. 9, 3, and 6) of in-
  creasing length at the time of 1 per second. Then, they are asked
to repeat to the tester the numbers starting from the lowest
  value to the highest. The test is composed of 28 groups of
  numbers going from 2-figure as minimum to the maximum of
  8-figure numbers. Measures: number of correct answer.
- **Motor speed and coordination**, evaluated with a token motor
task. Patients are given 100 plastic token scattered in the way of
  not being overlapped, and they are asked to place them in a
  container as quickly as possible in a 60 s time, taking only 1 per
  hand and not making them slide to the table edge. Measure:
  number of tokens placed in the container during the first 30 s
and the final 30 s.

- **Processing speed**, evaluated with a symbol coding task. Patients receive a key explaining how unique symbols correspond to the individual numerals 1–9. They are asked to fill in the corresponding number beneath a series of symbols as quickly as possible. There is a 90 s time limit. Measure: number of correct items

- **Verbal fluency**, derived from Semantic Fluency and Letter Fluency. For Semantic Fluency, patients are given 60 seconds to name as many words as possible within a given category (i.e. “supermarket object”). Measure: number of words generated. For Letter Fluency, in two separate trials, patients are given 60 s to generate as many words as possible, starting with a specific letter (F and R in the version A, T and M in the version B). Measure: number of words generated.

- **Planning (executive functions)**, evaluated with the Tower of London. Patients are presented with two pictures simultaneously. Each picture shows three different colored balls arranged on three pegs, but the balls will be in a unique arrangement in each picture. The patient has to give the total number of times the balls in one picture would have to be moved in order to make the arrangement of balls identical to that of the other, opposing picture, not helping himself tracing passages with fingers and having 20 s for each picture. Measure: number of correct responses.

For each BACS subtest, both raw scores and Z scores are provided. Z scores are calculated based on healthy controls performance on the sample described in Anselmetti et al. (2008).

### 2.4. Anomaly detection task

Participants were presented with a total of 150 Italian sentences that included an equal number of well-formed sentences and sentences with violations, and were asked to decide if the sentences were correct. To minimize processing loads, sentences were presented visually, and the task was self-paced.

The sentences were subdivided in short (mean length: 33.38 characters) and long (mean length: 58.22 characters), to assess the impact of working memory requirements. Two sets were created, one including short and long sentences with syntactic violations and comparable correct sentences (syntactic set; 90 items in total), and one including sentences with semantic violations and comparable correct sentences (semantic set; 60 items in total). Presentation order was randomized within each set. Syntactic and semantic errors were constructed as follows:

- **Syntactic errors** were designed by relying on properties of Universal Grammar, with violations based on principle shared among all lexical items involved, rather than by a single lexical clash resulting from the computation of the entire sentence and a mismatch with respect to a specific syntactic or semantic interface. More specifically, three types of syntactic errors were designed, following a previous test used for frontotemporal dementia patients (Cotelli et al., 2007).

- Violations of locality principles with question formation (short label WhS), such as “Chi Giovanni vuole contattare l’infermiera prima di incontrare? (tr. *Who does John want to contact the nurse before meeting?). Locality principles refer to the restrictions on the possibility to interpret words that are “far” from the words they refer to (Manzini, 1992). For instance, an affirmative sentence like John wants to contact the nurse before meeting the doctor could yield two distinct interrogative sentences: Who does John want to contact before meeting the doctor versus *Who does John want to contact the nurse before meeting? The interrogative pronoun who is the complement of contact in the first sentence and the complement of meet in the second sentence. Despite the fact that the two sentences contain exactly the same number of words, only the first sentence is grammatically correct. In the second one, the syntactic relation between who and the verb it refers to (meet) is said not to be sufficiently “local” or, more specifically, to violate the so-called “Subjacency Principle”.

- Violations of locality principles with affirmative, specifically with clitic constructions (short label CIM), such as “Di queste foto, Maria ne pensa che Gianni vuole vedere due (tr. *Of these pictures, Maria of-them_plue thinks that Gianni wants to see two). In this case, the misplaced element is a clitic (i.e. stressless pronouns) which cannot move too far from the verb it refers to (compared the well-formed sentence Di queste foto Maria pensa che Gianni ne vuole vedere due, tr. *Of these pictures, Maria thinks that Gianni of-them_plue wants to see two).

- Wrong contrastive focus involving inversion (short label CFC), as in *Non arriva Gianni ma parte (tr. *Not arrives Gianni but leaves). Here the error is related to contrastive focus interpretation, where something is negated and contrasted with something else in the sentence. Focal interpretation, although it clearly has a dramatic impact at the level of discourse interpretation, is nevertheless syntactically implemented (Rizzi, 1997; Belletti, 1999). This operation is indeed sensitive to word order. From a given pair of grammatical sentence Gianni arriva (tr. Gianni arrives) and Arriva Gianni (tr. Arrives Gianni), the contrastive focus inversion can generate Gianni non arriva ma parte (tr. Gianni not arrives but leaves) and Non arriva Gianni ma Pietro (tr. Not arrives Gianni but Pietro). By contrast, *Non arriva Gianni ma parte (tr. *Not arrives Gianni but leaves) and *Gianni non arriva ma Pietro (tr. *Gianni not arrives but Pietro) are ungrammatical due to word order.

Semantic errors were based on semantic “contradiction” resulting from the computation of the entire sentence, as in Ho asciugato la maglietta con l’acqua (tr. I have dried my new shirt with water): the term “contradiction” here is used to distinguish the case of implausible semantic meaning which would result from situation which are possible but are nevertheless so rare to become impossible (such as *I have dried my new shirt with a volcano). This strategy of generating semantically anomalous sentences crucially differs from the standard one consisting in a lexical clash produced by a word which does not belong to the same semantic field as the others, as in I have dried my new shirt with an elevator which does not necessarily involve computation as it can be detected even locally as in I gave the dog a skyscraper. In the latter case, the anomaly is perceived when the subject reaches to the point of the sequence that contains the semantically unexpected word. In our test, the subject is rather asked to decide if the sentence is correct or incorrect on the basis of the global computation of all lexical items involved, rather than by a single “surprising” lexical item. Finally, our anomalies also differ from anomalies resulting from the computation of the entire sentence and a mismatch with respect to information in world knowledge, such as the classic “The Dutch trains are white” (Hagoort et al., 2004). In this case, the anomaly derives from computing the sentence meaning and detecting a mismatch with respect to a specific culturally driven world knowledge.

Throughout the syntactic and semantic set, interrogative and affirmative sentences with violations were used, and correct sentences had comparable sentence structure. Examples of items are provided in Table 1.

### 2.5. Statistical analyses

The STATISTICA Software for Windows, version 8 (StatSoft Inc., Tulsa, OK, USA) was used to perform the statistical analyses.

Differences between groups (patients vs controls) with respect to demographic and clinical variables, general intellectual ability, neuropsychological performances and anomaly detection task scores were analyzed. A chi-square test ($\chi^2$) was used to compare qualitative variables (i.e. gender), while t-tests were used to
compare quantitative variables (i.e. age, education, tests scores and task accuracy). The statistical significance level was set at $p < 0.05$, then adjusted according to Bonferroni correction when required.

The association between anomaly detection task scores and psychopathological features (only in patients), general intellectual ability and neuropsychological performances (in the two separate groups) was examined using standard tests of correlation (Pearson $r$). The statistical significance level was set at $p < 0.05$, then adjusted according to Bonferroni correction when required.

Since the experimental design for the anomaly detection task was factorial, single effects of each factor and their interactions were also analyzed. A general linear model (GLM), with diagnosis (two levels: patients vs controls), linguistic type (two levels: syntactic vs semantic), sentence type (two levels: correct vs anomalous) and length (two levels: short vs long) as fixed factors, was used. A more stringent significance level $p \leq 0.01$ was set to reduce the risk of type I errors that may accompany analytical comparisons. Post-hoc analyses with Tukey Test were performed for comparisons.

To further investigate the role of clinical and neuropsychological features on the anomaly detection task performances in patients with schizophrenia, we performed a separate GLM only in the patients sample. In detail, we focused on: duration of illness, as it is a variable generally associated to global cognitive impairment (Irani et al., 2011) and those scores that resulted significant in the previous correlation analysis (as will be shown, these are PANSS positive subscale, Tower of London and N-back). In this second GLM, accuracy rates were entered as dependent variable, accuracy rates were entered as dependent variable, duration of illness (years), PANSS positive subscale (score), Tower of London (score) and N-back performance (%2-back) as continuous predictors, while fixed factors were syntactic type (three levels: WHS vs CIM vs CFC) and sentence type (two levels: correct vs anomalous). The statistical significance level was set at $p < 0.01$ and then adjusted according to Bonferroni correction when required.

### 3. Results

#### 3.1. Demographic, clinical and neuropsychological features

Table 2 summarizes the demographic, clinical and neuropsychological data of patients with schizophrenia and healthy controls. No significant differences were found between patients and controls for demographic features. The antipsychotic treatment in the patients was distributed as follows: Forty-one patients were treated with clozapine (median daily dose 250 mg), eleven with haloperidol (median daily dose 3.5 mg), three with Risperidone (median daily dose 3 mg) and three with olanzapine (median daily dose 7.5 mg).

Regarding the neuropsychological measures evaluated both in patients and controls, the t-tests showed a significant difference in working memory performances. In detail, % of correct responses at both 1-Back and 2-Back conditions were significantly lower in patients with schizophrenia ($p \text{'}s \leq 0.0001$).

#### 3.2. Anomaly detection task

Table 3 summarizes, in healthy controls and patients, mean accuracy rates in the anomaly detection task, decomposed according to linguistic type (syntactic vs semantic), sentence type (correct vs anomalous) and length (short vs long).
Anomaly detection task performances, stratiﬁed by linguistic type and sentence type are shown in Fig. 1 (upper panel) and for the semantic set (lower panel), in patients (blue lines) and controls (red lines). (For interpretation of the references to color in this ﬁgure legend, the reader is referred to the web version of this article.)

3.3. Association between Anomaly detection task scores and psychopathological measures

Correlations tests between clinical measures and anomaly detection task performance in patients with schizophrenia, after applying Bonferroni correction, showed only one signiﬁcant correlation between the PANSS positive subscale score and the anomaly detection task in the case of long syntactically anomalous sentences ($r = -0.51, p = 0.001$), indicating that patients with more prominent positive symptoms showed lower accuracy rates for long syntactically anomalous sentences. No signiﬁcant correlations between clinical variables and the anomaly detection task in the semantic set were observed. Details are reported in Table 4.

In view of the signiﬁcant correlation between anomaly detection for long sentences in the syntactic set and the PANSS positive subscale, which includes the evaluation of disorganized thinking, and given the possible relation between language and thought disorder, we performed a correlation analysis also for PANSS P2 item “Conceptual Disorganization”. Results showed a signiﬁcant...
inverse correlation between Conceptual Disorganization scores and performance in the anomaly detection for long sentences of the syntactic type \( r = -0.61, p < 0.0001 \).

3.4. Association between Anomaly detection task scores and neuropsychological measures

Correlations (Bonferroni corrected) observed in patients between neuropsychological measures and anomaly detection task performances are shown in Table 5.

Accuracy rates correlated significantly with executive functions performances, more specifically planning abilities evaluated by means of Tower of London, in the case of syntactically anomalous sentences, both short \( r = 0.51, p = 0.002 \) and long \( r = 0.56, p = 0.001 \), and in the case of long semantically anomalous sentences \( r = 0.50, p = 0.002 \). Working memory abilities showed a significant effect only in the 2-back condition, in relation with accuracy rates for long semantically anomalous sentences \( r = 0.48, p = 0.003 \). The correlation analyses, Bonferroni corrected, in healthy controls are shown in Table 6.

Total IQ, measured with the TBI, was significantly correlated with accuracy rates for short and long syntactically anomalous sentences \( r = -0.52; r = 0.71, p = 0.01 \) and for short semantically anomalous sentences \( r = 0.51, p = 0.02 \). Total IQ was also negatively correlated with accuracy rates for short correct sentences in the syntactic set \( r = -0.53, p = 0.01 \). Working memory performances, in particular the % of correct responses in the 1-back condition, were significantly correlated with accuracy rates for short syntactically anomalous sentences \( r = 0.54, p = 0.01 \) and for short semantically anomalous sentences \( r = 0.52, p = 0.01 \).

3.5. General linear model analyses

Significant effects of single factors and interactions revealed by the general linear model (GLM), with diagnosis, linguistic type, sentence type and length as fixed factors and accuracy rates in the
anomaly detection task as dependent variable, are reported below. Significant effects were observed for:

- diagnosis ($F=33.5$, df = 1, $p < 0.0001$), with patients performing overall significantly worse than controls ($p < 0.0001$);
- linguistic type ($F=162.1$, df = 1, $p < 0.0001$), with overall accuracy in the syntactic set being significantly lower than accuracy in the semantic set, both in patients and controls ($p's < 0.0001$). Significant interactions were observed between:

- diagnosis*linguistic type ($F=22.90$, df = 1, $p < 0.0001$), with accuracy in the syntactic set being lower in patients compared to controls ($p < 0.0001$); moreover, accuracy in the syntactic set was lower than accuracy in the semantic set, both in patients and controls ($p's < 0.0001$).
- diagnosis*sentence type ($F=8.9$, df = 1, $p=0.003$), with accuracy on anomalous sentences being lower than accuracy on correct sentences within the patients group ($p < 0.0001$); moreover in patients accuracy on anomalous sentences was significantly lower than accuracy on both anomalous and correct sentences in controls ($p's < 0.0001$);
- diagnosis*linguistic type*sentence type ($F=8.3$, df = 1, $p=0.004$), with several significant differences (all $p's \leq 0.004$).

In summary, accuracy in the syntactic set was globally lower than accuracy in the semantic set, regardless of sentence type, and accuracy was globally lower in patients than in controls. Moreover, in patients the detection of syntactically anomalous sentences was significantly lower than in any other combination of factors.

No significant effects were observed for length.

In the subgroup of patients, the GLM with linguistic type, sentence type and length as fixed factors, duration of illness, PANSS positive sub scale, Tower of London and N-Back performances as continuous predictors, and accuracy rates on the anomaly detection task as dependent variable revealed significant effects of single factors and interactions, as reported below.

Significant main effects were observed for:

- Tower of London Total scores ($F=8.60$, df = 1, $p=0.004$), confirming a significant effect of executive functions on the accuracy rates on the anomaly detection task;
- linguistic type ($F=130.20$, df = 1, $p < 0.0001$), with overall accuracy in the syntactic set being significantly lower than accuracy in the semantic set ($p < 0.0001$);
- sentence type ($F=9.00$, df = 1, $p=0.003$), with overall accuracy in detecting anomalous sentences being significantly lower than accuracy in detecting correct sentences ($p=0.001$). A significant interaction was observed for:

- linguistic type*sentence type ($F=6.30$, df = 1, $p=0.01$), with accuracy on syntactically anomalous sentences being significantly lower than all other combinations (all $p's < 0.0004$).

Moreover, the accuracy on syntactically correct sentences was significantly lower than accuracy in judging sentences in the semantic set, both anomalous and correct ($p's < 0.0001$).

No significant effects were observed for duration of illness, PANSS positive subscale, N-Back performances, nor sentence length.

Considering only the syntactic set in the subgroup of patients, the GLM with syntactic type (three levels: WhS vs CIM vs CFC) and sentence type (two levels: correct vs anomalous) as fixed factors, duration of illness, PANSS positive sub scale, Tower of London and N-Back performances as continuous predictors, and accuracy rates on the anomaly detection task as dependent variable revealed significant effects of single factors and interactions, as reported below:

A significant main effect was observed only for:

- sentence type ($F=45.79$, df = 1, $p < 0.0001$), with overall accuracy in detecting anomalous sentences being significantly lower than for correct sentences ($p < 0.0001$).

No significant main effects were observed for duration of illness, PANSS positive subscale, Tower of London and N-Back performances, nor for syntactic type.

A significant interaction was observed between:

- syntactic type*sentence type ($F=4.74$, df = 2, $p=0.009$), with accuracy being significantly lower on anomalous sentences than correct sentences, independently of the syntactic type (all $p's < 0.003$). Among anomalous sentences, post-hoc analysis also showed a significant difference according to the syntactic type, with accuracy for WhS sentences being significantly higher than for CFC sentences ($p < 0.0001$), while no significant difference was observed between CIM and WhS nor CFC.

4. Discussion

This study was designed to investigate syntactic and semantic abilities of patients with schizophrenia as compared to healthy controls through the anomaly detection task. Stimuli of different length, containing syntactic and semantic violations, based respectively on Universal Grammar and compositional semantics, were used. Furthermore, the role of psychopathological and neuropsychological features was investigated through extensive correlation analysis.

The comparison of the accuracy rates in the anomaly detection task highlighted a significantly poorer performance in patients as compared to controls in the syntactic condition, but not in the semantic condition. When entering these data into a more global statistical analysis, the effects of diagnosis (patients vs controls) and linguistic type (syntax vs semantics) were confirmed. The difference in detecting syntactic versus semantic anomaly displayed by patients can be plausibly related to a specific impairment of syntactic knowledge, which selectively affects their ability in recognizing syntactic errors violating general principles of Universal Grammar. These results converge with previous studies showing deficits in the comprehension of sentences with complex syntactic structure with different experimental paradigms or in the context of a more comprehensive language assessment (Morice and McNicol, 1985; Lelekov et al., 2000; Bagnier et al., 2003; Tavano et al., 2008). Moreover, our data shed further light to previous evidence of reduced sensitivity in patients when confronted with syntactic anomalies either in terms of online judgments (Kuperberg et al., 1998) or in terms of integrative processes reflected in the P600 (Ruchshow et al. 2003; Kuperberg et al. 2006). With respect to this literature, our data provide a more precise focus on Universal Grammar rules as the core of the impairment at the syntactic level. Results are in agreement also with imaging evidence, which shows altered persylvian networks well-known for their involvement in syntax, including Broca’s area and the arcuate fascicle (Catani et al., 2011; Benetti et al., 2015).

As a further consideration, it is worth mentioning that, among the three types of syntactic anomalies considered here, violations of locality principles with question formation were less difficult to detect as compared to violations of locality principles involving clitic movement and wrong contrastive focus. A reasonable interpretation of this result is that the presence of a wh-element makes anomalies more evident for the parser to detect, and thus more
resistant to the deterioration of syntax in schizophrenia. The visibility of wh-phrases as opposed to the other type of movement is arguably due to the fact that wh-phrases activate an independent projection, while clitics are parasitic to an already existent head. A part from a previous work testing these three kinds of syntactic anomalies in demented patients and not mentioning differences across types (Cotelli et al., 2007), there is no literature on the topic. Further investigation is thus required to explore whether, among Universal Grammar rules, some are more vulnerable to impairment in pathological conditions.

As a second point, our results do not support the idea of a semantic deficit, as no significant differences were observed between patients and controls in the detection of semantic anomalies. A closer look at the data shows that the performance for patients and controls falls near the upper limit score, which might suggest that the semantic set was easier than the syntactic set and might raise sensitivity issues. However, performance did completely reach the ceiling, and the absence of differences between patients and controls is visible also in the most difficult conditions. In particular, there were no differences between patients and controls for anomalous long sentences, for which the lowest scores were obtained (92.27% and 93.56% accuracy for patients and controls, respectively). In all, although we cannot completely exclude sensitivity issues in the semantic task, our data do not offer evidence in favor of the hypothesis of a semantic impairment in schizophrenia, at least not in terms of semantic composition abilities. To this respect the literature is fragmented. While traditional literature conveys the general idea of impaired semantics, specific studies provide a more complex scenario. As discussed in the introduction, evidence of hyperpriming (Rossel and David, 2006) stands aside evidence of hypopriming (Barch et al., 1996). When tested with anomaly detections, there is evidence of insensitivity to semantic anomalies (Kuperberg et al., 1998) as well as evidence of normal N400 response (Ruchstow et al., 2003). Moreover, most of the studies reporting altered semantic associations are evaluated in the discourse context (Hella et al., 2013), rather than in sentential semantic composition. It is thus possible that semantic abnormalities in schizophrenia become visible when world knowledge and pragmatic aspects are involved, while sparing compositional abilities per se. Consistently with this hypothesis, in a preliminary study employing a novel protocol with 6 classic pragmatic tasks ranging from discourse organization to metaphor comprehension we reported a widespread communicative deficit in schizophrenia, possibly reflecting difficulties in integrating aspects of context such as previous discourse, world knowledge, speaker’s beliefs and intentions (Bambini et al. 2014; Bosia et al. 2015 (In Press)), which is in line with extensive yet fragmented literature in the field (Langdon et al. 2002; Brüne and Bodenstein, 2005; Champagne-Lavau and Stip, 2010).

From a more theoretical point of view, this study provides additional empirical evidence supporting the hypothesis that syntax and semantics are two autonomous linguistic levels of representation in the speaker’s mind/brain, in line with extensive evidence in the neurophysiological and neuroimaging literature (Moro et al., 2001; Cappa, 2012). Through the study of schizophrenia behavior, our test specifically revealed a dissociation between two core aspects of syntax and semantics, namely the recognition of Universal Grammar violations on the one hand and the recognition of semantic “contradictions” on the other hand. Crucially, syntactic and semantic capabilities are not contrasted here in terms of computational capacity vs. lexical interpretation, because the detection of semantic errors involved some type of formal computation as in the case of syntactic errors.

Another important aspect concerns the involvement of the psychopathological features and cognitive abilities on syntactic and semantic processing. With respect to clinical measures evaluated in patients, we found a significant correlation only with the PANSS positive subscale for long sentences in the syntactic set. The association with the PANSS positive score, which includes evaluation of disorganized thinking, suggests an influence of the former on linguistic performance, as previously hypothesized in the literature on the relation between language and thought disorder. This hypothesis was confirmed by a further analysis that specifically evaluated the correlation between accuracy in anomaly detection for long sentences in the syntactic set and PANSS P2 item “Conceptual Disorganization”. It is also possible that other positive symptoms assessed in the PANSS, especially delusions and hallucinations, which are known to interfere with sustained attention, may have to some extent hampered the recognition of syntactic anomalies in the case of long sentences. Still, no correlations were observed between the linguistic performance and any other psychopathological domains, including though disorders assessed with TLI, nor other putatively relevant factors, such as duration of illness. Importantly, when inserted as covariate in the global analysis, not even the PANSS Positive Subscale showed any effect on the anomaly detection task. To complete the discussion of psychopathology, we tested a homogeneous sample of clinically stabilized, treatment-responder patients with medium–low residual psychopathology. This allowed to minimize symptoms effects and thus to evaluate single cognitive and linguistic components in a more precise fashion. Yet we cannot exclude that symptomatology may directly affect linguistic performances, over a certain threshold.

Concerning the relationship with cognitive abilities, for patients we observed a correlation with planning abilities for the detection of anomalous sentences in the syntactic set (both short and long) and for long semantic sentences. A correlation with working memory (% 2-back) was also shown, but only for long semantic anomalies. When inserted as covariates in the global analysis, only planning abilities resulted to influence the task, while no effect of working memory was revealed. As a first consideration, the effect of planning abilities, present both in the syntactic and semantic conditions, may be a general feature of the task, as anomaly detection might require the formation of a strategy. Second, and more interestingly, working memory does not seem to influence the anomaly detection task. The correlation between working memory and the detection of long semantic anomalies, although significant, did not yield significant outcomes in the global analysis. Also, the correlation between the working memory score in the BACS and the detection of long syntactic anomalies, although with high r value (0.44), proved not significant after Bonferroni correction. Further evidence can be derived from the absence of significant effects of length. Throughout the set of syntactic and semantic stimuli, sentence length does not appear to be significant per se, since it is irrelevant for the error recognition, with patients making the same percentage of errors in the case of both long and short sentences. However, we cannot rule out that, beyond the syntactic deficit, working memory could have a slight influence on the task itself, when the sentences to process are longer, given the effects (although non significant) observed for long syntactic and semantic anomalies.

This point is especially relevant for syntax, given the extensive literature discussing the relation between sentence comprehension, and specifically syntactic processing, and working memory abilities (Just and Carpenter, 1992; Caplan and Waters, 1999, 2013; Lewis et al., 2006). The results of the present study indicate that an impairment in core syntactic knowledge, i.e. Universal Grammar, cannot be attributed to defective working memory as measured in the span task. Moreover, the deficit is present also when the requirements for working memory are minimized by the use of short sentences, which further supports the idea of a syntactic deficit independent of working memory load. In agreement with
Herrmann, M., 2008. cockpit failure:
neuropsychological dysfunction: disconnection in schiz...


